Scientific Journals
of the Maritime University of Szczecin

Zeszyty Naukowe
Akademii Morskiej w Szczecinie

2016, 46 (118), 101–107
ISSN 1733-8670 (Printed)
ISSN 2392-0378 (Online)
DOI: 10.17402/125

Received: 31.08.2015
Accepted: 04.04.2016
Published: 27.06.2016

An estimation method of message receiving probability for a satellite automatic identification system using a binomial distribution model

Takahiro Seta✉, Hiroshi Matsukura, Taro Aratani, Kenkichi Tamura

National Maritime Research Institute, Japan
Shinkawa 6-38-1, Mitaka, Tokyo, 181-0004, Japan
✉ corresponding author, e-mail: seta@nmri.go.jp

Key words: satellite AIS, recognition/detection probability, missing rate, coverage, statistical binomial distribution model

Abstract

Automatic identification system (AIS) data are used to analyze vessels' positions or maritime traffic. Recently, satellites are being adopted for gathering AIS data (satellite AIS). A satellite AIS can gather AIS data from all over the world, including the center of the ocean. However, because a satellite moves around the world every day, there is a very short window of time that the satellite can receive signals from some specific area. Furthermore, because a satellite AIS receives signals simultaneously from a wider area than a terrestrial AIS does, the satellite AIS has lower rate of signal reception, especially near vessel congested areas. This may cause many unrecognized vessels. For this situation, this paper proposes a new method to estimate the number of unrecognized vessels based on a binomial distribution model. With this method, we are able to estimate the number of unrecognized vessels just from satellite AIS data themselves.

Introduction

The automatic identification system (AIS) is a communication system for exchanging voyage or navigational information between vessels via radio waves. In addition, the information sent from vessels is also valuable for researchers and analysts on the land. Because AIS messages include the sender vessel's position, by gathering AIS messages we can know vessels' current positions and position histories. There are many services providing such AIS data (Shipfinder, 2013; Marine Traffic, 2015) and there also are many academic studies that could not have been done without AIS data.

Recently, satellites have become used for gathering AIS messages (satellite AIS). Satellites fly all over the world; therefore, we can gather vessels' positions all over the world. However, a satellite AIS has lower message receiving probability than a terrestrial AIS (AIS gathered by receivers on the coast). Because a satellite receives messages from much wider area simultaneously than a vessel does, interference among AIS messages from different vessels occur frequently. In particular, the message receiving probability becomes very low near vessel congested areas, and a satellite AIS fails to recognize many vessels in this situation.

For this situation, we need to know the message receiving probability of a satellite AIS, or how many unrecognized vessels the satellite AIS has. This is important to evaluate the potential of satellite AIS, and moreover, to refine satellite AIS data or to calculate magnification factor for statistics values made from satellite AIS data.

The simplest methods to estimate the vessel recognition rate of a satellite AIS are those by comparing with other data such as the terrestrial AIS data or some vessel list generated from other data sources, which can be treated as the correct data. These types of methods are used in, for example, ExactEarth (2011), Larsen et al. (2012). However, such methods can only estimate the vessel recognition rate for areas where we can gather vessel information in other ways than satellite AIS. Hence these methods are useful only for evaluating satellite AIS's potential, not for refining satellite AIS data. If we want to know information about areas where we can gather vessel position data only from satellite AIS, we need to develop some other method which does not use vessel data other than those from the satellite AIS. There are also methods which can estimate vessel recognition rate of a terrestrial AIS only from the gathered data themselves (Lapinski & Isenor, 2011; Hammond & Peters, 2012), however, those methods are designed for a terrestrial AIS because they implicitly assume a high message receiving probability or a long time span of gathering AIS data in the same area, both of which a satellite AIS does not have. There are also theoretical analyses of vessel detection probability for satellite AIS (Høye et al., 2008; Tunaley, 2011; Chen, 2014), however, these analyses are just theoretical models and do not appear to calculate the vessel detection probability from AIS data gathered in the real environment.

In this paper, we focus on the distribution of the counts of messages from individual vessels received in one satellite pass-over (Figure 1) and propose a new estimation method of message receiving probability for satellite AIS which is based on a binomial distribution model. This method is only applicable for satellite AIS data.

Problem

The main purpose of this paper is to develop an estimation method for message receiving probability or vessel detection probability for a satellite AIS which is applicable only using the gathered satellite AIS data themselves. The data we target to estimate these probabilities are the log of the raw AIS messages received by the satellite with timestamps.

The characteristics of a satellite AIS are as follows (Japan Aerospace Exploration Agency, 2015):

- Moves all over the world and gathers AIS data from all over the world.
- About 7000 m/s velocity and only 10~15 minutes to pass near one single vessel.
- About 1000 km elevation above the sea surface.
- About 1000 km radius footprint (sight).

Because of these characteristics, the vessel detection probability of a satellite AIS becomes low. Because a satellite passes in a short time, there are not many chances to receive AIS messages from each vessel. In addition, because a satellite is very fast, there is a Doppler shift effect. Because a satellite locates at very high altitude above the sea surface, the distance from vessels, which are at the sea surface, becomes far, resulting in weak signal strength. Because the footprint of sight is very wide, many ships can be in sight of a satellite simultaneously resulting very frequent signal collisions.

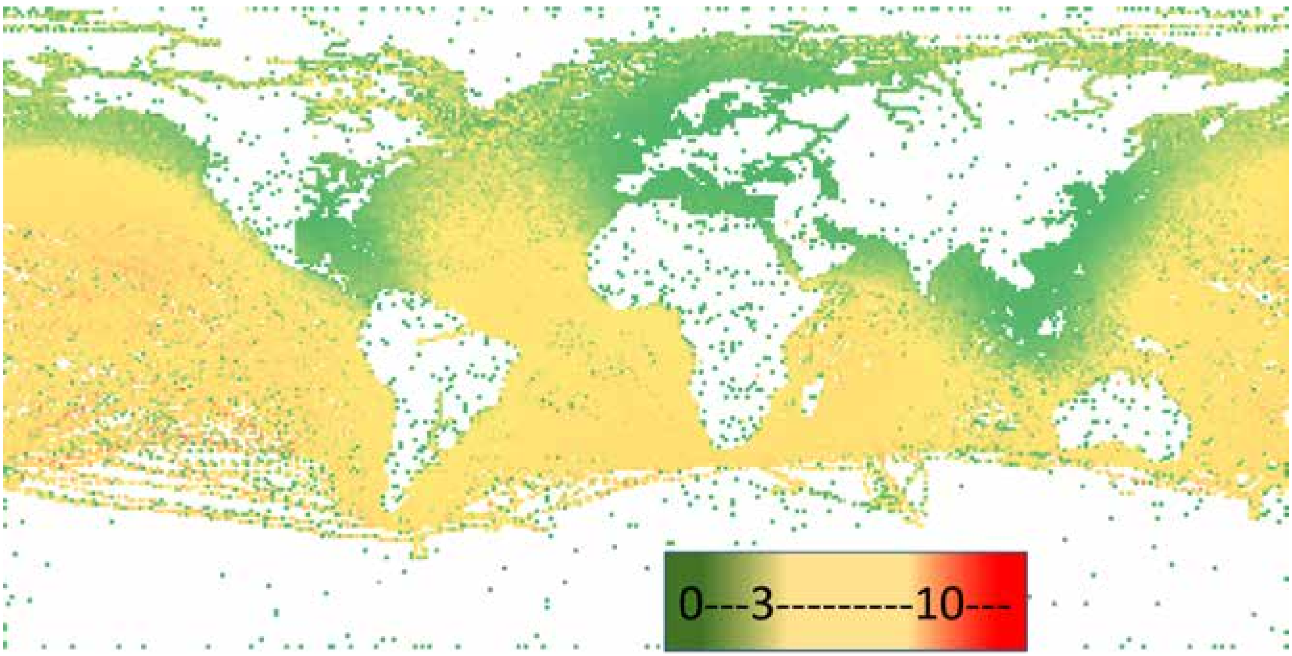

Figure 1. Average received message counts of satellite AIS from individual vessels in one pass-over

Method

In this section, we design an estimation method for the message receiving probability or the vessel detection probability for a satellite AIS.

Main idea

The main idea of our method is as follows:

- There may exist vessels that transmit AIS messages but are not recognized, because no transmitted message is received.
- Estimations or calculations in which the possibility of existence of unrecognized vessels is not mentioned may lead to incorrect results. True values and observed values are different.
- Because AIS messages are transmitted repeatedly, the average or the distribution of the received counts of AIS messages from each vessel should contain information on the receiving probability of AIS messages. Statistical models can extract this information.

The most important point is that this idea focuses on unrecognized/unobserved vessels. Existing methods use observed averages and other values, which are calculated only among received messages. Therefore, for example, average message receiving count for a single vessel is always larger than one. However, when there are unrecognized vessels, the true average message receiving count from one vessel can be less than one, because the message receiving counts of unrecognized vessels are zero. The difference of the true average and observed average is important. This difference becomes larger if the message receiving probability becomes lower, which is the case for satellite AIS, and most existing methods using observed averages will not work correctly for areas or data with low message receiving probability. A theoretical model for this difference will be shown in the following section.

Binomial distribution model

The binomial distribution model is a well-known statistical model. It is a discrete probability distribution model that assumes the following conditions.

- A trial comes to a result of "success" or "failure." The name "binomial" comes from this.
- Trials are repeated multiple times, n.
- Trials are independent of each other, and the "success" probability p of a single trial is the same for all repetitions (Bernoulli trial).

The distribution of the success count X from n repetitions with probability p is denoted as $B(n, p)$. The average success count (called the "mean" in statistics) and the probability with which the success count becomes k (called "probability mass function" in statistics) are shown in Table 1.

Table 1. Characteristics of binomial distribution model

Name	Value
Mean (average, expectation)	np
Probability mass function (the probability with which the success count becomes k, or $\Pr[X = k]$)	$\binom{n}{k} p^k (1-p)^{n-k}$

If we apply this model to the AIS message receiving situation, p represents the probability with which a single message is successfully received, and n represents the total number of messages transmitted from each vessel. Then, since $\Pr[X = 0] = (1 - p)^n$ represents the probability with which a vessel is not recognized, the average message receiving count of recognized vessels becomes $np/(1 - (1 - p)^n)$, which is the average without zero and which is larger than the true average of the binomial distribution model (Figure 2). If both p and n are small, $1 - (1 - p)^n$ becomes small and the difference between the true average and observed average becomes large. Therefore, analysis of data with low message receiving probability (small p) and gathered in a short time (small n), both of which are the case for satellite AIS, needs special treatment in which unrecognized vessels are mentioned.

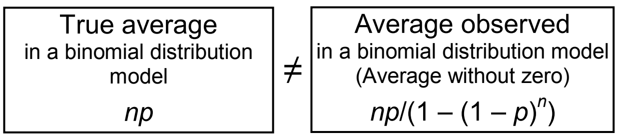

Figure 2. True average and observed average in binomial distribution model

The binomial distribution model may be too simple to express the physical phenomenon of AIS message loss (the success probability p should be different between vessels, and trials should not be independent – especially trials in near time), however it is easily analyzable and suitable for a first step. The result of a goodness-of-fit test is shown later in this paper.

Proposed estimation method

We propose the following estimation method in this paper.

1. Decide the target area to be analyzed.
2. Find the time period when the satellite pass by the target area. The time period should be long enough that all messages transmitted from the target area

and received by the satellite are included in the period. Thirty minutes will be enough since the satellite passes by in 10 minutes or so.

3. Decide n, which is the message transmitting count in one satellite pass-over. Because a satellite pass-over takes 10 minutes approximately and messages are transmitted almost every 10 seconds, $n = 60$ will be suitable. (However, from the experiments shown later in this paper, it appears that n is not so important if we only need to estimate vessel detection probability.)
4. Count the received messages from each vessel in the target area. We should also include messages from outside the target area if vessels move around both within the target area and outside, in order to avoid under-estimation.
5. By varying the message receiving probability p (or equivalently np), find the most suitable value p (or np) for which the theoretical average observed message count $np/(1-(1-p)^n)$ matches the average count of received messages.
6. From the estimated message receiving probability p, we can estimate the probability that a vessel remains unrecognized as $(1-p)^n$ and the true count of existing vessels as $1/(1-(1-p)^n)$ times larger than the observed count.

Note: Making a distribution by merging multiple periods is an interesting idea. However, because vessels do not stay in same area and n or p will be different among vessels, the distribution will become much less close to a binomial distribution. So, we chose a shorter time period, in which vessels can be assumed to stay in a small area with same message receiving probability p.

Experimental results

To evaluate the success of the proposed method, we show some experimental application results using some data near the Japan coastal area.

Sample data

In order to evaluate the proposed method, we use satellite AIS data gathered in September 2011 and terrestrial AIS data near the Japan coastal area in the same period. The terrestrial AIS data set is one of the most reliable data sets the authors have, which can be assumed to have almost 100% recognition rate for AIS equipped vessels along the Japan coastal area.

As the target area to analyze, we choose four areas near the Japan coast shown in Table 2 and Figure 3. These four areas are selected for the following reasons: (A) Tokyo-wan Bay is one of the most congested and important port areas in Japan. (B) Kii-suido Channel is one of the most important passing points in Japan; it is the entrance of Seto-naikai Inland Sea and Shio-no-misaki Cape is in the area. (C) Kanmon Area is a congested area itself and one of the areas that are close to China and Korea, both of which have congested ports. Because of the wide footprint of the satellite, it is challenging condition to have congested areas. (D) Tsugaru-kaikyo Strait is also an important passing point in Japan, but a little far from congested areas. The vessel recognition rates calculated by comparing satellite AIS data and terrestrial AIS data are as shown in Figure 4 and Table 3. The time periods of the data used are also in Table 3. The time period of area (C) is a little different to other areas, because no data was gathered in the area in the same time period as the other areas.

Table 2. Longitudes and latitudes of target areas

	Area Name	Longitude / Latitude
(A)	Tokyo-wan Bay	E139–E140 / N35–N36
(B)	Kii-suido Channel	E134–E136 / N33–N34
(C)	Kanmon Area	E130–E131 / N34–N35
(D)	Tsugaru-kaikyo Strait	E140–E141 / N41–N42

Figure 3. Locations of target areas

Table 3. Number of vessels detected by satellite AIS and terrestrial AIS

	Area Name	Time Period	Terrestrial	Satellite	Total
(A)	Tokyo-wan Bay	Sept. 15th 0:45–1:00 UTC	314 (100%)	156 (50%)	314
(B)	Kii-suido Channel	Sept. 15th 0:45–1:00 UTC	67 (87%)	45 (58%)	77
(C)	Kanmon Area	Sept. 14th 13:55–14:05 UTC	34 (100%)	6 (18%)	34
(D)	Tsugaru-kaikyo Strait	Sept. 15th 0:45–1:00 UTC	20 (87%)	19 (83%)	23

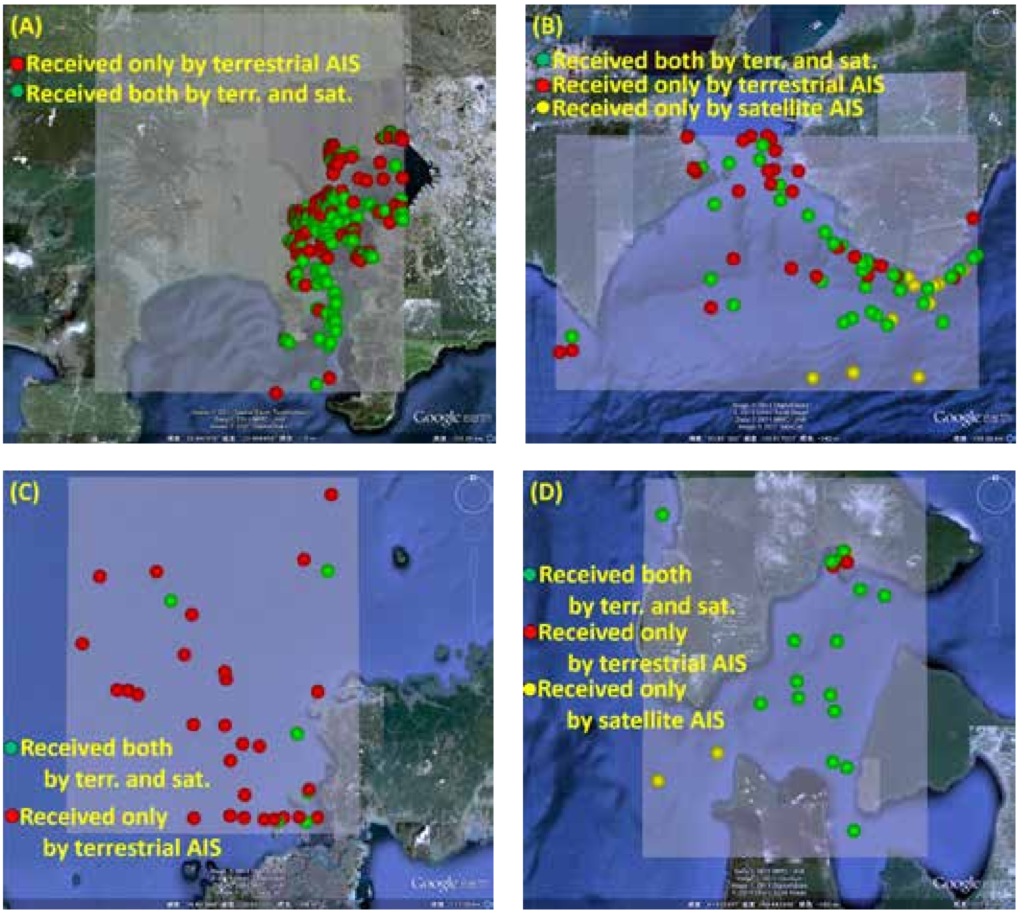

Figure 4. Comparison of detected vessels between satellite AIS and terrestrial AIS

Results and discussion

The proposed method is applied to the data shown above. The target areas and the time periods are already given. Though the time periods look shorter than those proposed, all AIS messages from the target areas are included. We set the message transmitting count $n = 60$ as suggested. Then, average receiving counts, suitable message receiving probability p, and other values are calculated as shown in Table 4. The comparison results between the proposed method and the simplest method, which is comparison against "correct" terrestrial AIS data, are shown in Table 5.

For the calculation shown in Table 4, we use np as the control value, rather than p, because np is closer in value to the observed average and easy to compare. As shown in Table 5, the trend of the proposed method and the simplest method matches well, though differences exist in area (A) and area (B). In areas where we can gather terrestrial AIS data, like the Japan coastal area, we can choose

Table 5. Comparison of estimated vessel detection probability

	Area Name	Proposed Method	Compare with Terrestrial AIS
(A)	Tokyo-wan Bay	78%	50%
(B)	Kii-suido Channel	69%	58%
(C)	Kanmon Area	14%	18%
(D)	Tsugaru-kaikyo Strait	85%	83%

Table 4. Calculation of vessel detection probability using proposed method

	Area Name	Obs. Avg.	Theory True Avg. np	p	Detection Prob. $1-(1-p)^n$	Theory Obs. Avg.
(A)	Tokyo-wan Bay	1.924	1.50	0.025	0.78	1.920
(B)	Kii-suido Channel	1.681	1.16	0.019	0.69	1.681
(C)	Kanmon Area	1.076	0.15	0.0025	0.14	1.076
(D)	Tsugaru-kaikyo Strait	2.206	1.88	0.031	0.85	2.207

Table 6. Vessel detection probability estimated from observed average received message count

n		Observed Average								
		1.001	1.01	1.1	1.5	2.0	2.5	3.0	5.0	10.0
30	np	0.002	0.02	0.20	0.90	1.62	2.26	2.85	4.98	10.0
	p	6.7e–5	6.6e–4	6.7e–3	0.03	0.054	0.075	0.095	0.166	0.333
	$1-(1-p)^n$	0.002	0.020	0.182	0.599	0.811	0.905	0.950	0.996	1.000
60	np	0.002	0.02	0.20	0.89	1.61	2.25	2.84	4.97	10.0
	p	3.3e–5	3.3e–4	3.3e–3	0.015	0.027	0.038	0.047	0.083	0.167
	$1-(1-p)^n$	0.002	0.020	0.181	0.592	0.804	0.899	0.945	0.994	1.000
104	np	0.002	0.02	0.20	0.88	1.59	2.23	2.82	4.97	10.0
	p	2.0e–7	2.0e–6	2.0e–5	8.8e–5	1.6e–4	2.2e–4	2.8e–4	5.0e–4	0.001
	$1-(1-p)^n$	0.002	0.020	0.181	0.585	0.796	0.892	0.940	0.993	1.000

the simplest comparison method; however, in the other areas, where satellite AIS plays a more important role, the only method we can apply is the proposed method. The results shown in Table 5 tell us that the proposed method is worth applying in such areas. Applying the proposed method, we can estimate how many undetected vessels are in the target area. When we do some statistical analysis using the satellite AIS data, we can use the estimated vessel detection probability as the magnification factor, which will refine the result of the statistical analysis. For example, in area (C), we can estimate there are almost 7.1 (= 100/14) times as many as observed vessels by the proposed method, where the most likely value is 5.5 (= 100/18) which is calculated from the simplest comparison method with 30% difference (7.1/5.5 = 1.3).

As another evaluation of the proposed method, we create a conversion table from observed average message counts to the corresponding vessel detection probabilities for various message transmitting counts n as shown in Table 6. Using this kind of table, we can easily estimate the true vessel count in the target area from the observed vessel count and average received message count.

In Table 6, we create the conversion table with three different transmitting counts n, which are 30, 60 and 10 000. Of course, 10 000 messages are never transmitted from one vessel in one satellite pass-over, and 30 messages are far smaller than the suggested suitable value 60. However, as shown in Table 6, the vessel detection probabilities $1-(1-p)^n$ are not very different between these three values. Therefore, we can say that the value of n is not very significant in the proposed method if we only need the vessel detection probability, and we do not need to worry about finding the true value of n.

Goodness-of-fit test

As mentioned above in the method section, the binomial distribution might be too simple to express the physical phenomenon of AIS message loss. However, it will be useful to test the goodness-of-fit between the observed distributions and the theoretical distributions. For this purpose, we applied the χ^2 equality test.

The test is done as follows, using statistical computing software ("*R*"):

1. Create the observed distribution of message counts.
2. Create the theoretical distribution from a binomial distribution, and remove the probability for zero, which will never be observed. (*dbinom* function of *R*).
3. Apply χ^2 test to the above distributions and get the p-value. (*chisq* test function of *R*).

The result of the χ^2 equality test is shown in Table 7. As shown in the table, the p-values are around 0.40, which means that almost the same amount of differences appear with probability of 40% even if two datasets are come from exactly same distribution. Therefore, we cannot deny the hypothesis that the distribution of the observed data follows a binomial distribution. (n = 120 of the Kanmon Area says that the compared distributions are different, but this is because n = 120 is too high.)

Table 7. Results of goodness-of-fit test

	Area Name	n = 30	n = 60	n = 120
(A)	Tokyo-wan Bay	0.3306	0.3787	0.4137
(B)	Kii-suido Channel	0.3453	0.3895	0.4215
(C)	Kanmon Area	0.4031	0.4314	0.001861
(D)	Tsugaru-kaikyo Strait	0.3306	0.3787	0.4137

The p-values are shown. N.B. (A) and (D) coincidentally resulted in the same value (not a copying mistake).

Conclusions

In this paper, we proposed a new estimation method for message receiving probability and vessel detection probability for satellite AIS. The method is based on the binomial distribution model and only requires the target data to be analyzed. Furthermore, we also created a conversion table from average received message counts to vessel detection probabilities using the proposed method. These make the proposed method easy to apply. The estimation results are also evaluated using well-known coastal area data. Though the method may need some improvement, the trend of estimation results is good. Using the proposed method, we can calculate vessel detection probabilities of any area, which can be used as magnification factors in statistical analysis, resulting in refined analysis results. Therefore, we can conclude that the proposed method will be a very useful tool, when we use satellite AIS data.

We may also conclude that the message receiving count is important. Some AIS data services provide the latest message of each vessel only; however, if they add message receiving count of each vessel in some time span to the provided data, we may extract much more information from the data.

As an area for further study, it would be interesting to use defined transmission interval by the rule (International Telecommunication Union, 2014). Though we conclude that transmitting count *n* is not important, it is not denied that we can improve the estimation result using more detailed estimation of transmission counts.

Acknowledgments

The satellite AIS data used in this paper are served from ExactEarth. The terrestrial AIS data along the Japan coast area used in this paper are served from TST Corporation (Toyo Shingo Tsushinsha). Mr. Kawabe, Mr. Shibata and other members from IHI Corporation and IMC Co., Ltd. gave useful comments and information especially on the characteristics of satellites and current trends. The authors express their sincere gratitude to all of them.

References

1. Chen, Y. (2014) Detection Probability Analysis of Space-Based AIS Signal by Using Geometrical Model. *Journal of Shipping and Ocean Engineering* 4. pp. 161–171.
2. ExactEarth (2011) exactNews, issue 1.
3. Hammond, T.R. & Peters, D.J. (2012) Estimating AIS Coverage from Received Transmissions. *The Journal of Navigation* 65 (3). pp. 409–425.
4. Høye, G.K., Eriksen, T., Meland, B.J. & Narheim, B.T. (2008) Space-based AIS for global maritime traffic monitoring. *Acta Astronautica* 62. pp. 240–245.
5. International Telecommunication Union (2014) *Technical Characteristics for an Automatic Identification System Using Time Division Multiple Access in the VHF Maritime Mobile Frequency Band*. Recommendation ITU-R M.1371-5.
6. Japan Aerospace Exploration Agency (2015) SPAISE – SPace based AIS Experiment (in Japanese) [Online] Available from: http://www.satnavi.jaxa.jp/spaise/index.html [Accessed: May 15, 2016]
7. Lapinski, A.-L.S. & Isenor, A.W. (2011) Estimating Reception Coverage Characteristics of AIS. *The Journal of Navigation* 64 (4). pp. 609–623.
8. Larsen, J.A., Nielsen, J.F.D., Mortensen, H.P., Rasmussen, U.W., Laursen, T. & Ledet-Pedersen, J. (2012) Evaluation of AIS Reception in Arctic Regions From Space by Using a Stratospheric Balloon Flight. *Polar Record* 48 (1), pp. 39–47.
9. Marine Traffic (2015) www.marinetraffic.com.
10. Shipfinder (2013) www.shipfinder.com.
11. Tunaley, J.K.E. (2011) *The Performance of a Space-Based AIS System*. London Research and Development Corporation Technical Report.